Gu Test: A Progressive Measurement Of Generic Artificial Intelligence

Author: Lifan Gu□□□□□

Turing Test is invalid. It is subjective, just an empirical test, not a scientific experiment. Langauge conversations do not have a full coverage of human intelligence.

Sciences are different from mathematics. Scientific experiements only can falsify, but never prove unlimited possiblities. Scientific research is an ongoing process, should always open to new experiments.

So other existing empirical tests for AI technologies, such as the regular Go games played by AlphaGo Zero and other computer Go systems, the simulations and road tests of self-driving cars, the datasets for natural langauge understanding, etc. are also inadequate.

Technological Singularity is baseless. Driverless cars with no constrains (i.e. SAE level 5 automated-driving) are impossible. There are problems in the definition of SAE level 4.

In reality, there is no way to prove a car with SAE level 4 automated-driving ability, especially when the mode evolution in future is not stable. So new concepts of AI and new definitions of automated-driving should be studied.

In this paper, I will discuss the problems in Turing Test, the problems in existing testing of AlphoGo Zero, self-driving cars, natural langauge understanding, and the problems in the mainstream textbook AI: A Modern Approach. Then I will propose Gu Test, a progressive measurement of generic artificial intelligence, based on falsifiability, which could help to develop scientific intelligence theories gradually.

1. The Problems in Turing Test

Turing Test is invalid, but still cause misleading widely in AI research so far.

Many existing tests for AI techonologies have similar problems as Turing Test. So it is important to analyze its problems and clarify the misleading.

Turing Test is subjective. Testing it with different people could yield very different results. People with different knowledges, especially with different understanding levels of computer technologies, could give very different results. The subjectiveness of Turing Test cause unstable results, which makes Turing Test invalid.

Moreove, langauge conversations do not have a full coverage of human intelligence. So Turing Test can not test the non-language parts of human intelligence. Indistinguishablity between humans and computers by language conversations does not mean equivalence of intelligence.

Turing Test is also an empirical test, not a scientific experiment.

Sciences are different from mathematics. Scientific experiements only can falsify, but can never prove unlimited possiblities. Actually, equivalence of intelligence between humans and computers can never be proved, but only can be falsified.

Scientific research is an ongoing process, should always open to new experiments. If computers pass some tests, other people still could design new tests to disprove.

Scientific experiments should be done with strictly controlled conditions, to test the underlying principles. Scientific conclusions can only be derived from these principles based on the strict conditions. From empirical tests, people can not derive scientific conclusion.

Other existing tests for AI technologies have many similar problems. In the next sections, I will discuss the testing problems for computer Go systems, self-driving cars, and natural langauge understanding